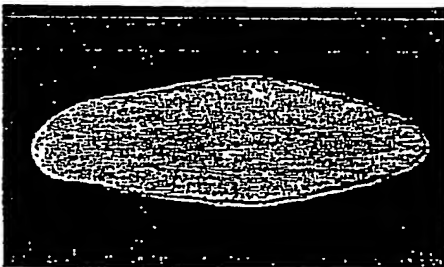


FIG.1

File Edit View Search Go Bookmarks Tasks Help

EST ID: D21.6.Contig
Accession: AY062374
cDNA Sequence: CTTCTGCAAGGTCCACAGTTACCATTAAACAAATCGAGTCCTGCAGTAAAGGAATATGAAA
TGAGACAATCCTACAATTTTCTGGAGCACCTATGGGGGATCAGTTCAAATTCATAGCA
ATGTGCTTCTGCTGTAGAGGCGAGAAAGAGAAAAGAGAGAAATGCNAGATCTTAATG
AAAGGCTAGCTAATTATATTGAAAAGGTAAGATTTCTAGAAGCTCNAACAAAAGATTAA
CAAATGAATTGAATACGTTACGTGAAAGATGGGTNAAGAAGCTGAAACGATACGAGCTT
TATATGAGATTGAAATGGATCAATTGAAAAAGTTATTAGACGAAGCTGAAGCTGCTAGAT

Category: Cytoskeletal/Structural
Subcategory: intermediate filaments
In Situ: 

Query= D21.6.Contig (1828 letters) Database: nr 775,385 sequences; 246,391,666 total letters Searching done

Sequences producing significant alignments:	Score (bits)	E Value
dbj BAB64909.1 (AB056671) intermediate filament b [Dugesia...	471	e-131
emb CAA68255.1 (X99996) intermediate filament protein [Lin...	412	e-114
emb CAB38180.1 (AJ004937) cytoplasmic intermediate filamen...	411	e-114
emb CAA69027.1 (Y07747) IF protein [Lineus viridis]	397	e-109
gb AAD29248.1 AF101065.1 (AF101065) intermediate filament g...	392	e-108

>dbj|BAB64909.1| (AB056871) intermediate filament b [Dugesia japonica]
>dbj|BAB64910.1| (AB056872) intermediate filament b [Dugesia japonica]
Length = 610

Score = 471 bits (1211), Expect = e-131
Identities = 244/567 (43%), Positives = 365/567 (64%), Gaps = 10/567 (1%)
Frame = +3

Query: 27 NXSSPAVKEYEMRQSYNFSGAPMGGSVQIHSNVSSAVEGREEREREMXDLNERLANYIEK 206
N+++ ++ EM++SY S P S +HS V++ + GRE+EK E+ +LN+R ANYI+K
Sbjct: 45 NQNASSIRTLEMKKSYSVSATPGCATSNIVHSGVNNLMNGREKEKNEQLQELNDRFANYIDK 104

Query: 207 VRFLEAXNKRLTNELNTRLRERWGXEAERIRALYEIEMDQLKKLLDEAFAARSELLPKINK 366
VR LE NKRLT+ELN L++WG E RI+ALY+ +H QL++ LD+AEA++++L KIN
Sbjct: 105 VRSLEDENKRLTDELNDLKQWGNETARIKALYDSMSQLRRSLDQAEASKAQLEMKINT 164


-FIG. 2-

SUBSTITUTE SHEET (RULE 26)

File Edit View Go Favorites Tools Window Help

The Sánchez Laboratory [home](#) | [SmedDb](#) | [search](#) | [contact us](#) | [researcher's login](#)
 REGENERATION RESEARCH

- [Home](#)
- [Search](#)
- [SmedDb](#)
- [Search VIA Blast](#)
- [Publications](#)
- [Regeneration In the News](#)
- [Lab Members](#)
- [Contact Us](#)
- [Credits](#)
- [Other Info](#)



Search *Schmidtea mediterranea* EST Database
 Frequency Distribution | SmedDb Flowchart | Functional Distribution

Please check off a Category and/or Sub-categories to search, or type in an Accession Number:

Search via Accession Number:


<input type="checkbox"/> Cell defense (32)	<input type="checkbox"/> CELL-cell communication (136)
<input type="checkbox"/> carrier proteins (2)	<input type="checkbox"/> adhesion (27)
<input type="checkbox"/> DNA repair (2)	<input type="checkbox"/> extracellular matrix (33)
<input type="checkbox"/> immunology (?) (0)	<input type="checkbox"/> hormone/growth factors/activators (19)
<input type="checkbox"/> stress response (26)	<input type="checkbox"/> other membrane proteins (34)
	<input type="checkbox"/> receptors (24)
<input type="checkbox"/> Cytoskeletal/Structural (139)	<input type="checkbox"/> DNA Replication/Modification (130)
<input type="checkbox"/> intermediate filaments (10)	<input type="checkbox"/> apoptosis (31)
<input type="checkbox"/> microfilaments (47)	<input type="checkbox"/> cell cycle/division (56)
<input type="checkbox"/> microtubules (35)	<input type="checkbox"/> chromosome/nuclear structure (21)
<input type="checkbox"/> others (27)	<input type="checkbox"/> DNA synthesis (22)
<input type="checkbox"/> thick/thin filaments (20)	
<input type="checkbox"/> General Metabolism (228)	<input type="checkbox"/> Intracellular Signaling (211)
<input type="checkbox"/> amino acids (25)	<input type="checkbox"/> channels/transporters (39)
<input type="checkbox"/> lipids (66)	<input type="checkbox"/> effectors/modulators (39)
<input type="checkbox"/> nucleotides (18)	<input type="checkbox"/> protein kinases (44)
<input type="checkbox"/> other enzymes/cofactors (56)	<input type="checkbox"/> protein phosphatases (16)
<input type="checkbox"/> sugar/glycolysis (65)	<input type="checkbox"/> transduction (37)
<input type="checkbox"/> Mitochondria (72)	<input type="checkbox"/> No Match (1281)
<input type="checkbox"/> Protein Metabolism (175)	<input type="checkbox"/> RNA Metabolism (137)
<input type="checkbox"/> degradation (63)	<input type="checkbox"/> RNA binding (17)
<input type="checkbox"/> folding (23)	<input type="checkbox"/> RNA polymerases (2)
<input type="checkbox"/> post-translational modification (10)	<input type="checkbox"/> RNA processing (44)
<input type="checkbox"/> ribosomal components (39)	<input type="checkbox"/> transcription factors (77)
<input type="checkbox"/> translation factors (32)	
<input type="checkbox"/> tRNA synthesis (10)	<input type="checkbox"/> Unknown Function (357)
<input type="checkbox"/> Secretory Pathways (88)	<input type="checkbox"/> invertebrates (175)
<input type="checkbox"/> endoplasmic reticulum (27)	<input type="checkbox"/> plants (12)
<input type="checkbox"/> golgi (28)	<input type="checkbox"/> Prokaryotes (9)
<input type="checkbox"/> nuclear components (33)	<input type="checkbox"/> vertebrates (159)
	<input type="checkbox"/> yeast (4)

internet zone

File Edit View Go Favorites Tools Window Help

The Sánchez Laboratory [home](#) | [SmedDb search](#) | [contact us](#) | [researcher's login](#)
REGENERATION RESEARCH

- Home
- Search SmedDb
- Search VIA Blast
- Publications
- Regeneration In the News
- Lab Members
- Contact Us
- Credits
- Other Info



Search *Schmidtea mediterranea* EST Database
Frequency Distribution | SmedDb Flowchart | Functional Distribution

Please check off a Category and/or Sub-categories to search, or type in an Accession Number:

Search via Accession Number:

<input type="checkbox"/> Cell defense (32)	<input type="checkbox"/> CELL-cell communication (136)
<input type="checkbox"/> carrier proteins (2)	<input type="checkbox"/> adhesion (27)
<input type="checkbox"/> DNA repair (2)	<input type="checkbox"/> extracellular matrix (33)
<input type="checkbox"/> immunology (?) (0)	<input type="checkbox"/> hormone/growth factors/activators (19)
<input type="checkbox"/> stress response (26)	<input type="checkbox"/> other membrane proteins (34)
<input type="checkbox"/> Cytoskeletal/Structural (139)	<input type="checkbox"/> receptors (24)
<input type="checkbox"/> intermediate filaments (10)	<input type="checkbox"/> DNA Replication/Modification (130)
<input type="checkbox"/> microfilaments (47)	<input type="checkbox"/> apoptosis (31)
<input type="checkbox"/> microtubules (35)	<input type="checkbox"/> cell cycle/division (56)
<input type="checkbox"/> others (27)	<input type="checkbox"/> chromosome/nuclear structure (21)
<input type="checkbox"/> thick/thin filaments (20)	<input type="checkbox"/> DNA synthesis (22)
<input type="checkbox"/> General Metabolism (228)	<input type="checkbox"/> Intracellular Signaling (211)
<input type="checkbox"/> amino acids (25)	<input type="checkbox"/> channels/transporters (39)
<input type="checkbox"/> lipids (66)	<input type="checkbox"/> effectors/modulators (39)
<input type="checkbox"/> nucleotides (18)	<input type="checkbox"/> protein kinases (44)
<input type="checkbox"/> other enzymes/cofactors (56)	<input type="checkbox"/> protein phosphatases (16)
<input type="checkbox"/> sugar/glycolysis (65)	<input type="checkbox"/> transduction (37)
<input type="checkbox"/> Mitochondria (72)	<input type="checkbox"/> No Match (1281)
<input type="checkbox"/> Protein Metabolism (175)	<input type="checkbox"/> RNA Metabolism (137)
<input type="checkbox"/> degradation (63)	<input type="checkbox"/> RNA binding (17)
<input type="checkbox"/> folding (23)	<input type="checkbox"/> RNA polymerases (2)
<input type="checkbox"/> post-translational modification (10)	<input type="checkbox"/> RNA processing (44)
<input type="checkbox"/> ribosomal components (39)	<input type="checkbox"/> transcription factors (77)
<input type="checkbox"/> translation factors (32)	<input type="checkbox"/> Unknown Function (357)
<input type="checkbox"/> tRNA synthesis (10)	<input type="checkbox"/> invertebrates (175)
<input type="checkbox"/> Secretory Pathways (88)	<input type="checkbox"/> plants (12)
<input type="checkbox"/> endoplasmic reticulum (27)	<input type="checkbox"/> Prokaryotes (9)
<input type="checkbox"/> golgi (28)	<input type="checkbox"/> vertebrates (159)
<input type="checkbox"/> vesicular components (33)	<input type="checkbox"/> yeast (4)

Search via blast

internet zone

F1662

**** Sanchez Laboratory ** REGENERATION RESEARCH - Netscope 6**

The Sánchez Laboratory
REGENERATION RESEARCH

[Home](#) [Search](#) [SmedDb](#) [Search VIA](#) [Blast](#) [Publications](#) [Regeneration](#) [In the News](#) [Lab Members](#) [Contact Us](#) [Credits](#) [Other Info](#)

[home](#) | [SmedDb search](#) | [contact us](#) | [researcher's login](#)

Schmidtea mediterranea EST Database Results
Your search produced 10 results

EST ID Accession Number	Description	Category Sub-category	Whole Mount InSitu
D.21.5.Contig AY068374 More Info	(AB056871) intermediate filament b [Dogesia-jaborica]	Cytoskeletal/Structural * intermediate filaments	view InSitu
H.102.6e AY068720 More Info	(AF339450) hilarin [Hrudo medicinalis]	Cytoskeletal/Structural * intermediate filaments	(none available)
H.102.6e(T3) AY068721 More Info	(AF339450) hilarin [Hrudo medicinalis]	Cytoskeletal/Structural * intermediate filaments	(none available)
H.106.2E AY067086 More Info	(AB056871) intermediate filament b [Dogesia-jaborica]	Cytoskeletal/Structural * intermediate filaments	(none available)
H.12.11A AY067087 More Info	(AB056871) intermediate filament b [Dogesia-jaborica]	Cytoskeletal/Structural * intermediate filaments	view InSitu
H.24.5b AY068722 More Info	(Y07747) IF protein [Lineus viridis]	Cytoskeletal/Structural * intermediate filaments	(none available)

FIG.4

The Sánchez Laboratory
REGENERATION RESEARCH

- [Home](#)
- [Search](#)
- [SmedDb](#)
- [Search VIA](#)
- [Blast](#)
- [Publications](#)
- [Regeneration](#)
- [In the News](#)
- [Lab Members](#)
- [Contact Us](#)
- [Credits](#)
- [Other Info](#)

Blast Search

Search name

Blast Parameters

Search type

Expected Value

Sequence

```
CTTCTGCAAGGTCCACAGTTACCATTAAACAATCGAGTCCTGCAGTAAAGGAATATGAAA  
TGAGACAATCCTACAAATTTTCTGGAGCACCTATGGGGGATCAGTTCAAATTCATAGCCA  
ATGTCTCTTCGCTGTAGAGGCCGAGAAAGAGAAAAGAGAGAAATGCGNAGATCTTAATG  
AAGGCTAGCTAAATTATATGAAAAGGTAGATTTCTAGAAGCTCNAAACAAAGATTAA
```

** Sanchez Laboratory ** REGENERATION RESEARCH - Netscape 6

FIG. 5A

BLASTM 2.1.2 [Nov-13-2000]

Reference:

Allscnui, Stephen F., Thomas L. Madden, Alejandra A. Schffer,
Jinghui Zhang, Zheng Zhang, Webb miller, and David Lipman (1997),
"Gapped BLAST and PSI-BLAST: a new generation of protein database search
programs", Nucleic Acids Res. 25:3389-3402.

Query TEST

(1828 letters)

Database: Smed

3890 sequences; 2,526,228 total letters

Sequences producing significant alignments:

Score E
(bits) Value

ref Smed4531 D21.6.Contig	2495	0.0
ref Smed3960 HB.24.7g	36	0.057
ref Smed2705 H.97.6h	36	0.057
ref Smed2761 H.9.7f	34	0.23
ref Smed3456 H.111.3b	34	0.23
ref Smed2084 H.83.9h	34	0.23
ref Smed1429 H.44.7g	34	0.23
ref Smed3687 H.44.2c(T3)	32	0.89
ref Smed3653 H.86.9d	32	0.89
ref Smed3328 H.2.7a	32	0.89
ref Smed3058 H.50.4e	32	0.89
ref Smed1991 H.57.2c	32	0.89
ref Smed1880 H.59.12c	32	0.89
ref Smed448 H.17.3e	30	0.89
ref Smed4748 HB.9.4f	30	3.5
ref Smed4734 H.67.11d(T3)	30	3.5
ref Smed4692 H.73.7c	30	3.5
ref Smed4308 HB.35.4e	30	3.5
ref Smed4263 HB.29.8e	30	3.5
ref Smed4244 HB.22.6a	30	3.5
ref Smed3859 H.98.8c(T3)	30	3.5

FIG. 5B-1

ref Smed3806 H.29.9d(T3)	30	3.5
ref Smed3670 H.40.7b(T3)	30	3.5
ref Smed3284 H.29.10f	30	3.5
ref Smed3224 H.30.1e	30	3.5
ref Smed3143 H.38.10f	30	3.5
ref Smed2842 H.9.4b	30	3.5
ref Smed2143 H.35.6a(T3)	30	3.5
ref Smed1718 H.49.1d(T3)	30	3.5
ref Smed1992 H.6.1e	30	3.5
ref Smed1332 H.104.5d	30	3.5
ref Smed1208 H.83.7g	30	3.5
ref Smed1167 H.24.6d	30	3.5
ref Smed476 H.92.5f	30	3.5
ref Smed464 H.69.2a	30	3.5
>ref Smed4531 D21.6.Contig		
Length = 1828		
Score = 2495 bits (1258), Expect = 0.0		
Identities = 1281/1281 (100%)		
Strand = Plus / Plus		
Query: 1	cttctgcaagggtccacagttaccatttaocaaatcgagtcctgcagtaagggaatatgaaa	60
Sbjct: 1	cttctgcaagggtccacagttaccatttaocaaatcgagtcctgcagtaagggaatatgaaa	60
Query: 61	tgagacaatcctacaatttttctggagcacctatgggggatcagttcaaattcatagca	120
Sbjct: 61	tgagacaatcctacaatttttctggagcacctatgggggatcagttcaaattcatagca	120

FIG.5B-2

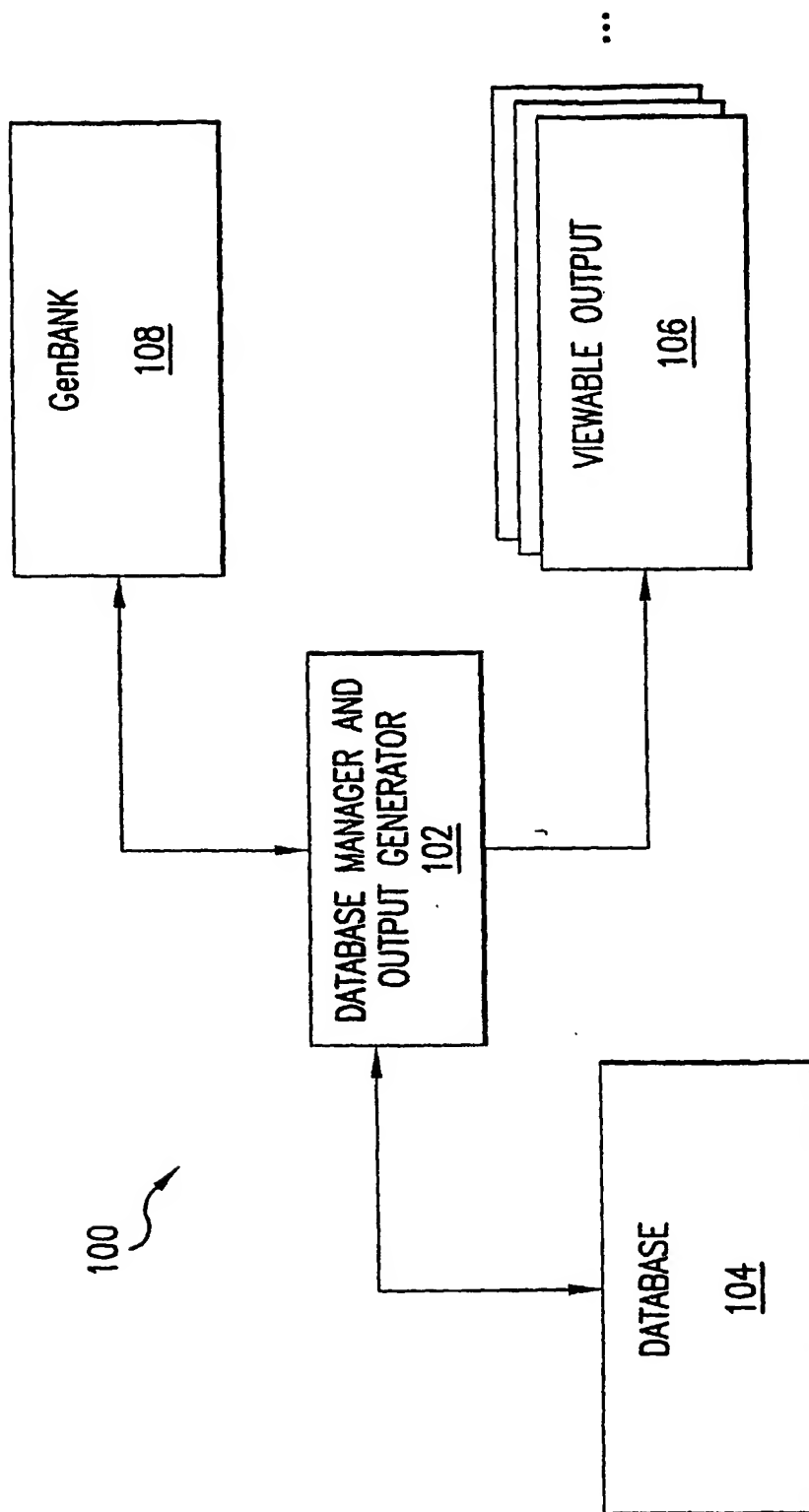


FIG.6

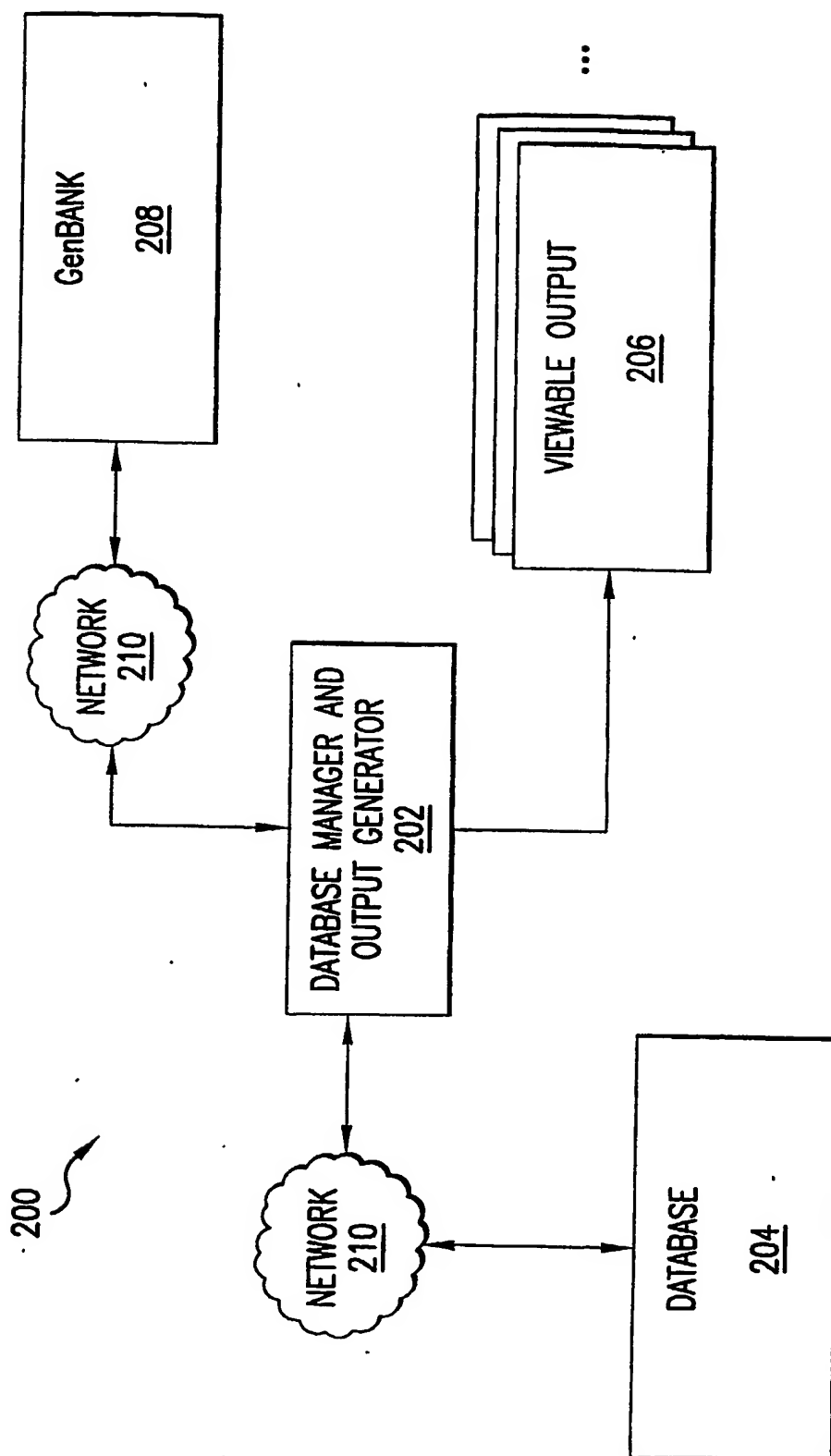


FIG. 7

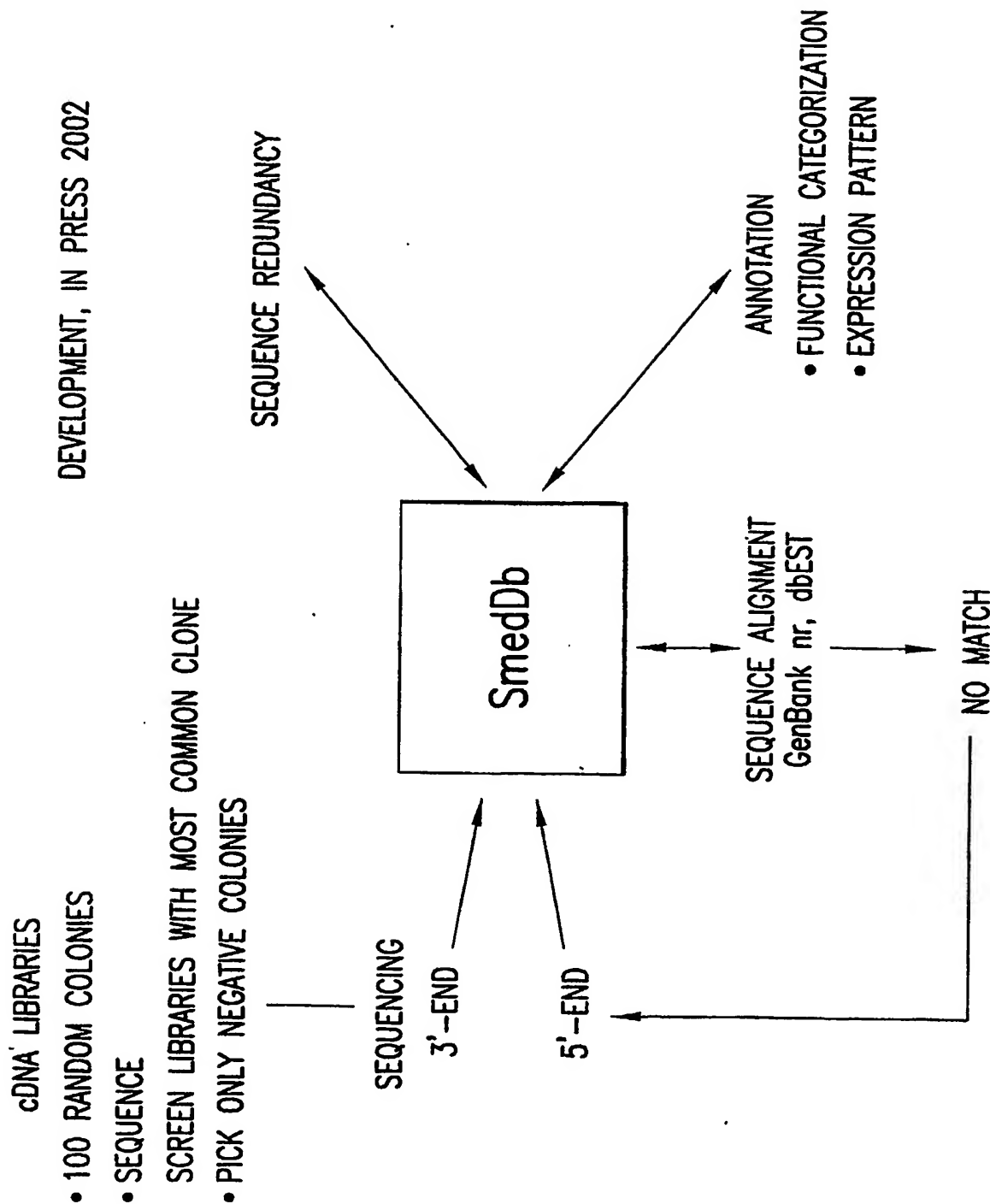
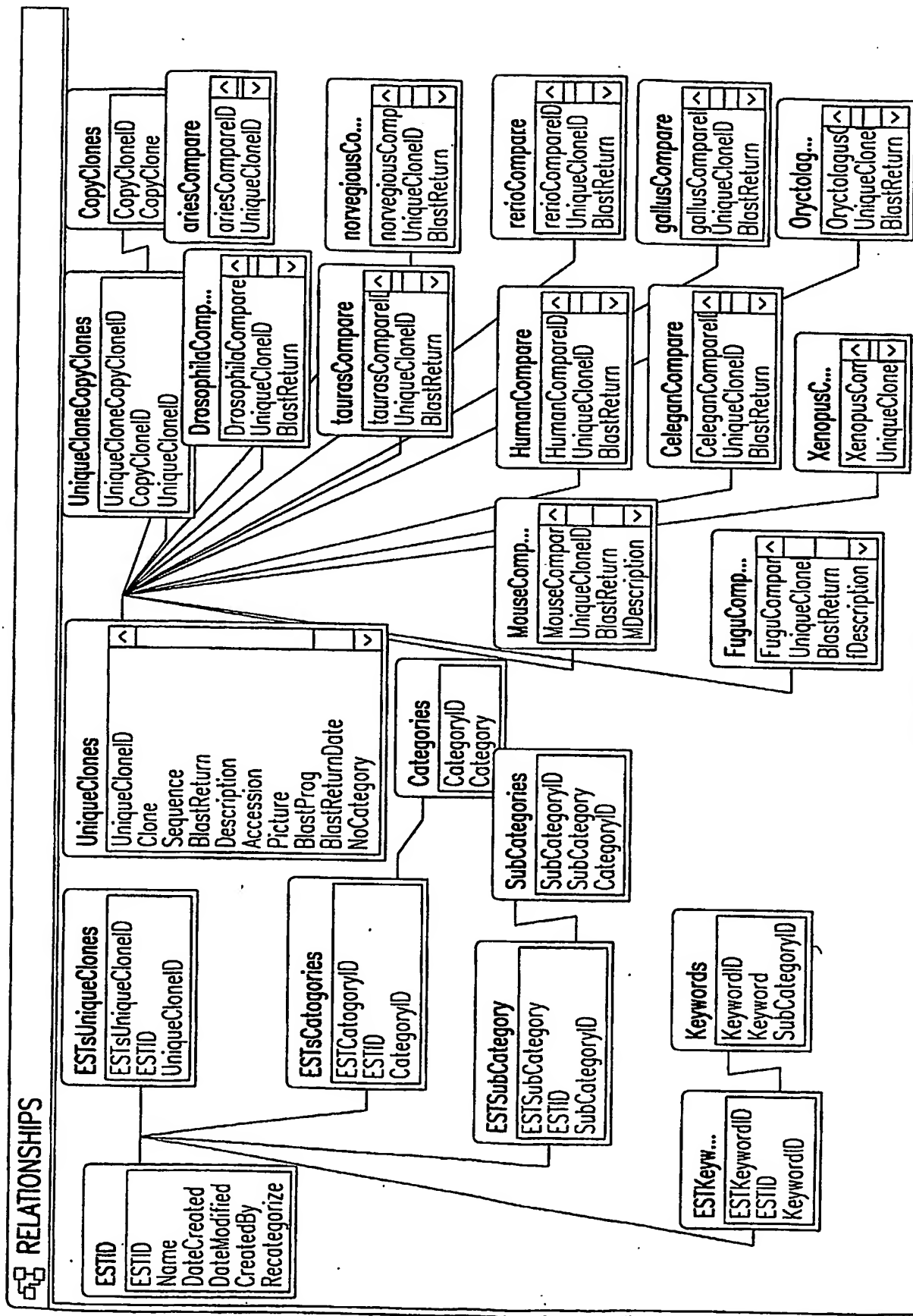


FIG.8



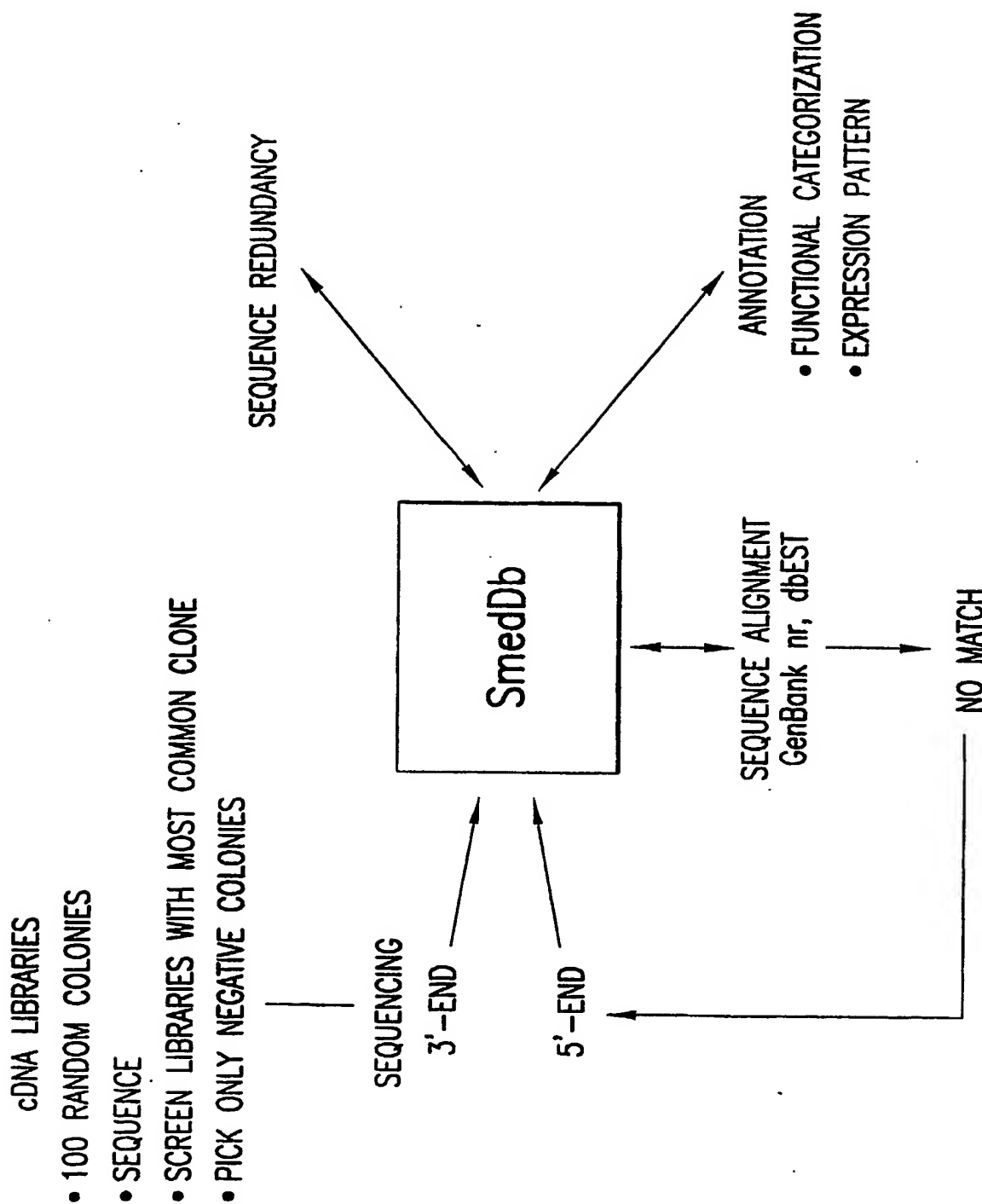


FIG.10A

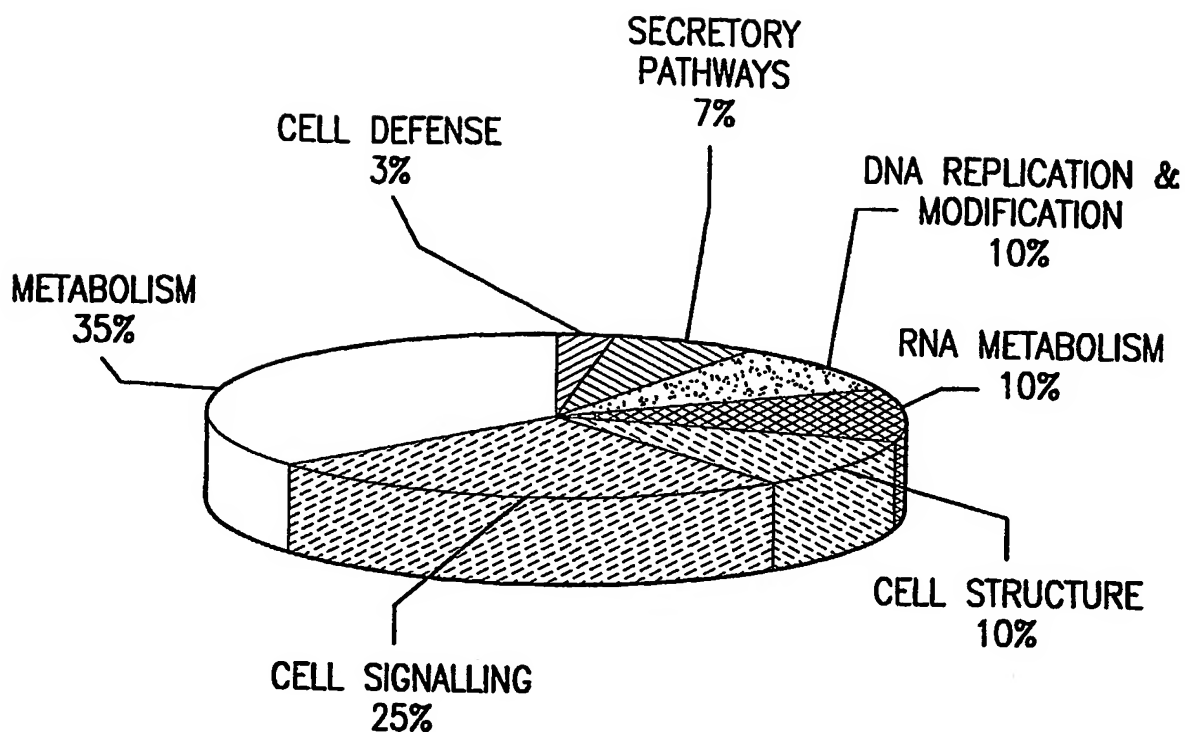


FIG.10B

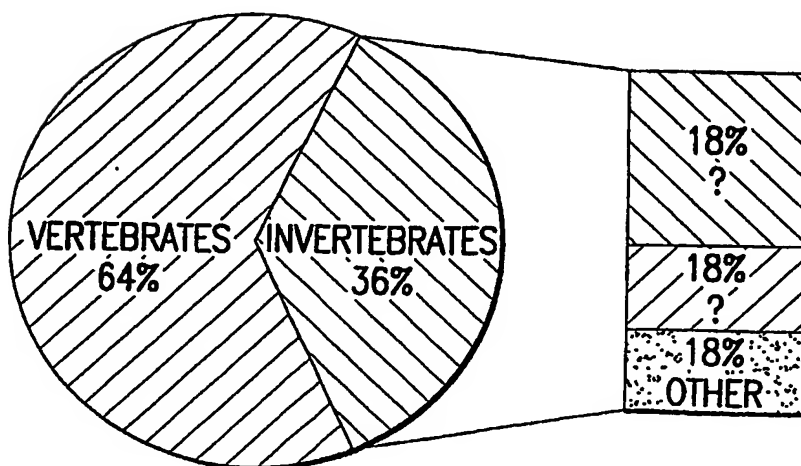


FIG.10C

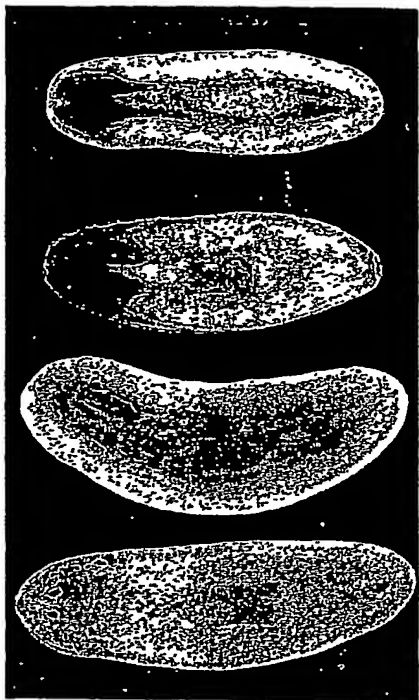


FIG. 11A

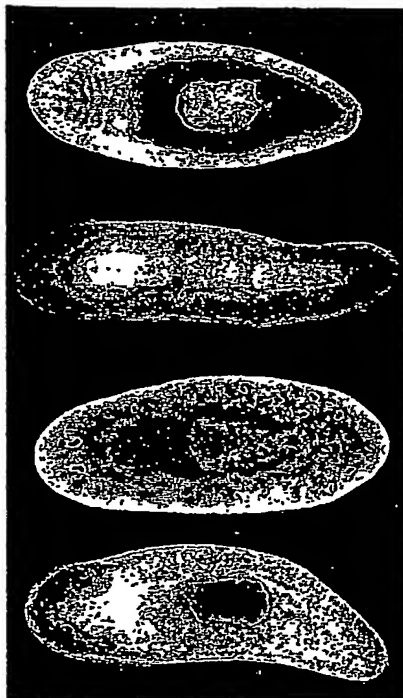


FIG. 11B

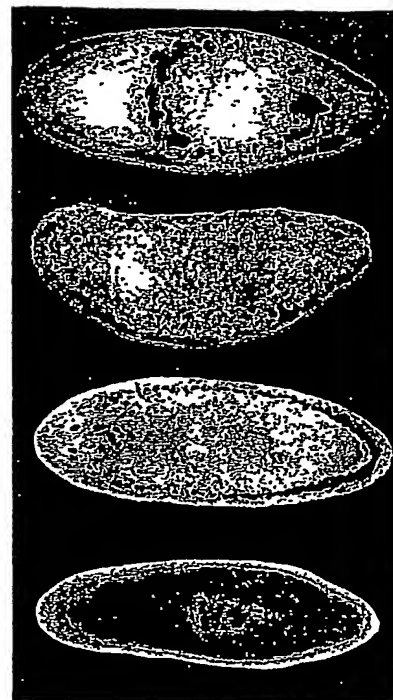


FIG. 11C

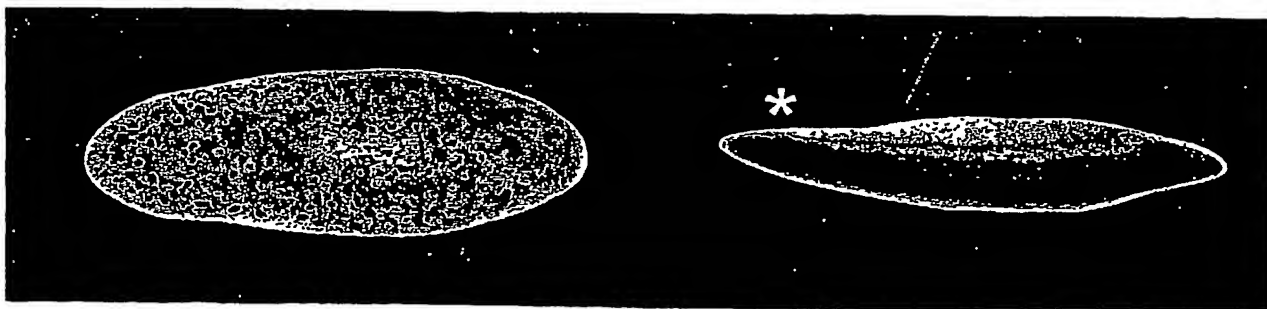


FIG. 11D

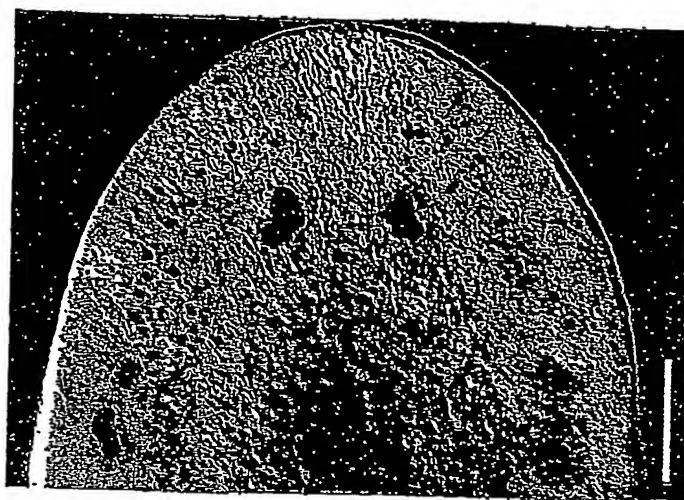


FIG.12A

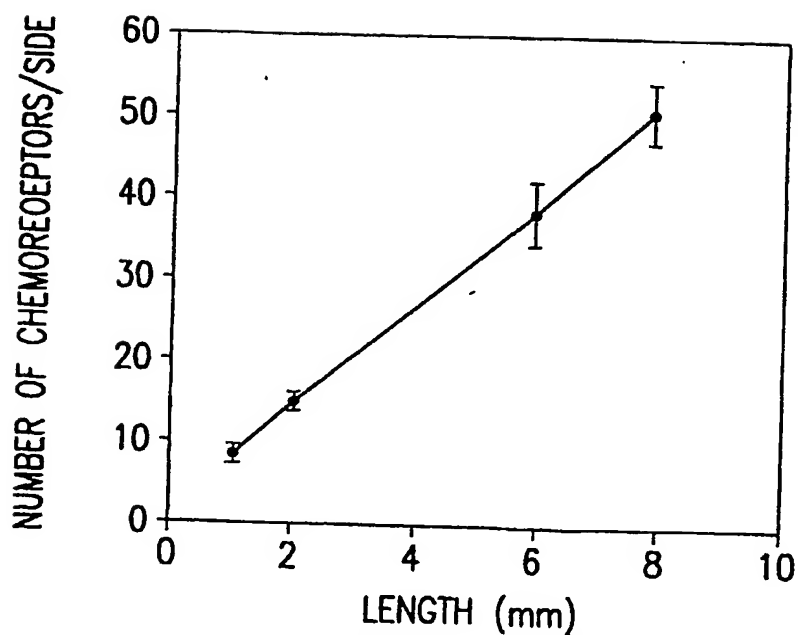


FIG.12B

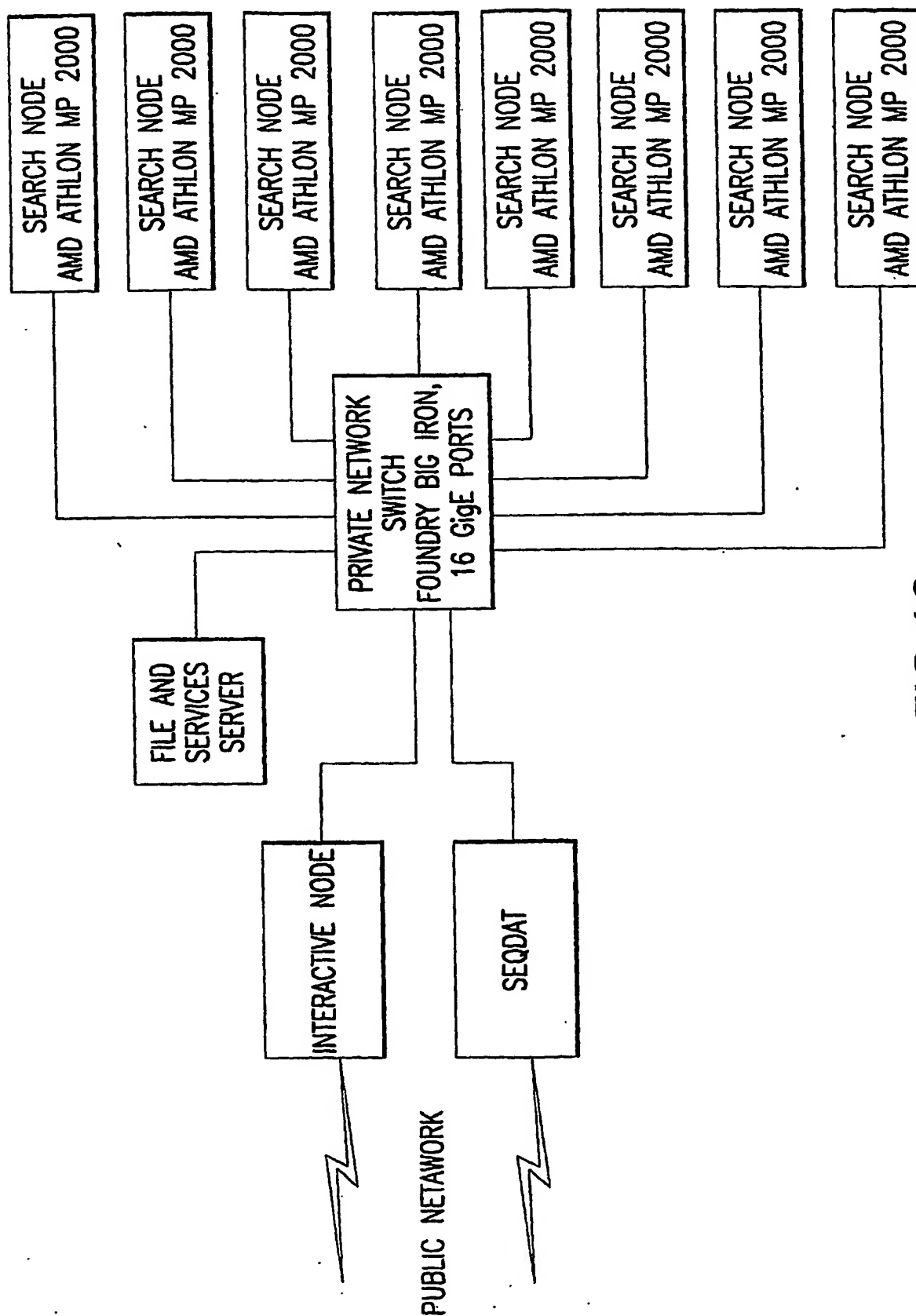


FIG.13

```
#!/usr/sbin/perl
#
#check if this is a first time search on the input files
#
print "Is this a first time search on datafiles in \${HOME}/newsearch (reply yes or no): ";
chomp($flag1 = <STDIN>);

if ($flag1 eq 'yes') {
$direct = newsearch;
}
#
#if not a first time search
#
if ($flag1 eq 'no') {
$direct = search;
}
#
#check if new db available - if no then exit
```

FIG.14A

```

#
$max = @dirs[0];
foreach $n (@dirs) {
    if ($n gt $max)
        { $max = $n; }
}
#
# Now, $max is the latest date-coded directory
#
$Dir_ID = $max;

$date_last = qx (grep Posted /uufs/incc.utah.edu/common/home/planaria/$direct/$Dir_ID/1x12.htm | uniq | cut -d "
" -f 8,9,10);
#
#Again change format
# Turn date_last into something useful
#
($m, $d, $y) = split(/,?/, $date_last);
if ($m eq 'Jan')
    { $month = '01'; }
if ($m eq 'Feb')
    { $month = '02'; }
if ($m eq 'Mar')
    { $month = '03'; }
if ($m eq 'Apr')
    { $month = '04'; }
if ($m eq 'May')
    { $month = '05'; }
if ($m eq 'Jun')
    { $month = '06'; }
if ($m eq 'Jul')
    { $month = '07'; }
if ($m eq 'Aug')
    { $month = '08'; }
if ($m eq 'Sep')
    { $month = '09'; }
if ($m eq 'Oct')
    { $month = '10'; }
if ($m eq 'Nov')
    { $month = '11'; }
if ($m eq 'Dec')
    { $month = '12'; }

$d =~ s/,//;
if ($d < 10)
    { $d = '0' . $d; }

chomp ($y);

$date_last_useful = "$y-$month-$d";
print "Date of last database: ", $date_db_useful, "\n";

print "Date of last search db: ", $date_last_useful, "\n";
#
#Compare two dates-if first not later than second exit with message
#

```

FIG. 14 (Cont'd)

```

if (date_db_useful le $date_last_useful)
{die "No database update since last search. Try again tomorrow as updates are checked for nightly\n";}
}
#
#check if both searches or only first to be done
#
print "Do you want to do the first search only (enter 1) or fist search and tblastx search (enter 2): ";
chomp($flag = <STDIN>);
print "The information that follows are the jobids for the batch jobs submitted to complete the search.\n";
print"\n";
#
#do first stage of search
#
#figure out how many nodes are free to use for search
#
$busy = qx(showq|grep Nodes|cut -d " " -f26);
$active = qx(showq|grep Nodes|cut -d " " -f31);
$noes = $active - $busy;
if ($nodes == 0)
{
    $nodes = 1;
}
print "NODE TO BE USED: ", $nodes, "\n";
#
#make necessary directory for output of search
#
($DAY, $MONTH, $YEAR) = (localtime)[3,4,5];
$MONTH++;
if ($MONTH < 10)
{
    $MONTH = '0' . $MONTH;
}
if ($DAY < 10)
{
    $DAY = '0' . $DAY;
}

```

FIG. 14C-1

```

$YEARS = $YEAR + 1900;
if ($flag eq 1)
{system("mkdir -p \${HOME}/${direct}/${YEARS}-${MONTH}-${DAY}");}
else
{system("mkdir -p \${HOME}/${direct}/${YEARS}-${MONTH}-${DAY}/TBLASTX");}
#
#set up scripts for first search
#
$source = "${ENV{HOME}}/${direct}/searchlist.in";
open(IN, " $source");
@line = <IN>;
foreach (@line)
{chomp;}
#
#Write out PBS header to each of the SCRIPT files being written
#
foreach (1..$nodes){
$name = "${ENV{HOME}}/${direct}/SCRIPT-S_";
open ($name, ">$name");
print $name ("##PBS -S /bin/csh\n");
print $name ("##PBS -q sequence\n");
print $name ("##PBS -W qos=2\n");
print $name ("##PBS -l walltime=500:00:00,nodes=1:ppn=2\n");
print $name ("##PBS -N Blast", $_, "\n");
print $name ("##PBS -e blast", $_, ".stderr\n");
print $name ("##PBS -o blast", $_, ".stdout\n");
}

```

FIG.14C-2

```

print $name ("setenv BLAST_CLI_HOME /uufs/sequence/sys/pkg/blast/std\n");
print $name ("setenv BLAST_DB /scratch/local/sofia/db\n");
print $name ("setenv WORKDIR $HOME/$direct\n");
print $name ("cd $WORKDIR/$YEARS,$MONTH,$DAY,$DAY,\"n\");
print $name ("ln -s $BLAST_CLI_HOME/.ncbirc .ncbirc\n");
print $name ("rm -r /scratch/local/sofia/db\n");
print $name ("mkdir -p /scratch/local/sofia/db\n");
print $name ("cp /uufs/sequence/sys/pkg/blast_db/std/nr.* /scratch/local/sofia/db/.n");
print $name ("cp /uufs/sequence/sys/pkg/blast_db/std/nt.* /scratch/local/sofia/db/.n");
print $name ("cp /uufs/sequence/sys/pkg/blast_db/std/sts.* /scratch/local/sofia/db/.n");
print $name ("cp /uufs/sequence/sys/pkg/blast_db/std/est.* /scratch/local/sofia/db/.n");
print $name ("cp /uufs/sequence/sys/pkg/blast_db/std/gss.* /scratch/local/sofia/db/.n");
print $name ("cp /uufs/sequence/sys/pkg/blast_db/std/htgs.* /scratch/local/sofia/db/.n");
}
$X = 1;
#
# Loop through input sequence file name in the file searchlist.in and generate
# the actual part of the PBS scripts that does the searches
#
foreach $line (@line) {
    $name = "$ENV{HOME}/$direct/SCRIPT-$X";
    if ($line =~ m#x#)
        {print $name ("\"$BLAST_CLI_HOME/blastall -a 2 -p blastx -e 1 -b 5 -v 5 -d \"$BLAST_DB/nr -i ../$line -o
$line.htm -T T \"n");
        if ($flag eq 2)
            {print $name ("perl /uufs/sequence/sys/pkg/blast/tblast_single.pl $line\n");
            print $name ("cd TBLASTX\n");
            $line =~ s#x##;
            print $name ("\"$BLAST_CLI_HOME/blastall -a 2 -p tblastx -e 1 -b 5 -v 5 -d \"$BLAST_DB/nt
$BLAST_DB/sts $BLAST_DB/est $BLAST_DB/gss $BLAST_DB/htgs\" -i ../$line -o $line.htm -T T \"n");
            print $name ("cd ..n");
        }
    }
    if ($line =~ m#n#)
        {print $name ("\"$BLAST_CLI_HOME/blastall -a 2 -p blastn -e 1 -b 5 -v 5 -d \"$BLAST_DB/nt
$BLAST_DB/sts $BLAST_DB/est $BLAST_DB/gss $BLAST_DB/htgs\" -i ../$line -o $line.htm -T T \"n");}
    $X++;
    if ($X > $nodes)
        { $X = 1; }
}
#
#submit search jobs
#
foreach (1..$nodes){
    $name = "$ENV{HOME}/$direct/SCRIPT-$_";
    close ($name, ">$name");
    qx(qsub $name);
}
#
# Print message to user that scripts have been successfully submitted
#
print "All batch jobs have been submitted. Check periodically by logging into sequence and doing a showq. Output
files will be found in $ENV{HOME}/$direct/$YEARS-$MONTH-$DAY for the first stage search and
$ENV{HOME}/$direct/$YEARS-$MONTH-$DAY/TBLASTX for the second stage search\n";

```

FIG. 14 (Cont'd)

```

#first get date of current database and store in date_db
#
qx{ln -s /uufs/sequence/sys/pkg/blast/std/.ncbirc .ncbirc};
$date_db = qx(/uufs/sequence/sys/pkg/blast/std/fastacmd -d /uufs/sequence/sys/pkg/blast/std/nr -I|grep Date|
cut -d " " -f5,6,7);
#
# Turn date_db into something useful
#
($m, $d, $y) = split(/ ?/, $date_db);
if ($m eq 'Jan')
    {$month = '01';}
if ($m eq 'Feb')
    {$month = '02';}
if ($m eq 'Mar')
    {$month = '03';}
if ($m eq 'Apr')
    {$month = '04';}
if ($m eq 'May')
    {$month = '05';}
if ($m eq 'Jun')
    {$month = '06';}
if ($m eq 'Jul')
    {$month = '07';}
if ($m eq 'Aug')
    {$month = '08';}
if ($m eq 'Sep')
    {$month = '09';}
if ($m eq 'Oct')
    {$month = '10';}

```

FIG.14E-1

```

if ($m eq 'Nov')
    {$month = '11';}
if ($m eq 'Dec')
    {$month = '12';}

$d =~ s/././;
if ($d < 10)
    {$d = '0' . $d}

chomp ($y);

$date_db_useful = "$y-$month-$d";
#
# get date of database used in last search and store in date_last
# parse date format to find newest directory

# @files is a list of all files in the search directory
# @dirs is a list of only the date directories in that search directory
#
@files = qx(ls #ENV{HOME}/$direct);
foreach (@files)
    {chomp;}
foreach (@files) {
    if (m#\d+\d+\d+#)
        {push @dirs, $_;}
}
#
# Find the greatest string in @dirs

```

FIG.14E-2


```

#PBS -S /bin/csh
#PBS -q sequence
#PBS -W qos=2
#PBS -l walltime=222:00:00,nodes=1:ppn=2
#PBS -N Blast2
#PBS -e blast2.stderr
#PBS -o blast2.stdout
#
# set all environment variables needed
#
setenv BLAST_CLI_HOME /uufs/sequence/sys/pkg/blast/std
setenv BLAST_DB /scratch/local/sofia/db
setenv WORKDIR $HOME/seq_dev_2/search
#
#move to proper directory
#
cd $WORKDIR/2003-07-16
#
#link needed so Blast works
#
ln -s $BLAST_CLI_HOME/.ncbirc .ncbirc
#
#Clean up local scratch from last run and get local copy of all
#necessary databases
#
rm -r /scratch/local/sofia/db
mkdir -p /scratch/local/sofia/db
cp /uufs/sequence/sys/pkg/blast_db/std/nr.* /scratch/local/sofia/db.
cp /uufs/sequence/sys/pkg/blast_db/std/nt.* /scratch/local/sofia/db.

```

FIG.15A

```

cp /uufs/sequence/sys/pkg/blast_db/std/stds.* /scratch/local/sofia/db.
cp /uufs/sequence/sys/pkg/blast_db/std/est.* /scratch/local/sofia/db.
cp /uufs/sequence/sys/pkg/blast_db/std/gss.* /scratch/local/sofia/db.
cp /uufs/sequence/sys/pkg/blast_db/std/htgs.* /scratch/local/sofia/db.
#
#The rest of the script runs the searches for the given input files (2x12, 9x12, and 16x12 shown)
#This particular script does both the first and second stage searches on each file;
#between the two stages the output of the first stage is parsed to check for searches
#in which "No hits found"; for these searches the second stage input files are then created.
#
$BLAST_CLI_HOME/blastall -a 2 -p blastx -e 1 -b 5 -v 5 -d $BLAST_DB/nr -i ../2x12 -o 2x12.htm -T T
perl /uufs/sequence/sys/blast/tblast_single.pl 2x12
cd TBLASTX
$BLAST_CLI_HOME/blastall -a 2 -p tblastx -e 1 -b 5 -v 5 -d "$BLAST_DB/nt $BLAST_DB/stds $BLAST_DB/est
$BLAST_DB/gss $BLAST_DB/htgs" -i ../2x12 -o 2x12.htm -T T
cd ..
$BLAST_CLI_HOME/blastall -a 2 -p blastx -e 1 -b 5 -v 5 -d $BLAST_DB/nr -i ../9x12 -o 9x12.htm -T T
perl /uufs/sequence/sys/blast/tblast_single.pl 9x12
cd TBLASTX
$BLAST_CLI_HOME/blastall -a 2 -p tblastx -e 1 -b 5 -v 5 -d "$BLAST_DB/nt $BLAST_DB/stds $BLAST_DB/est
$BLAST_DB/gss $BLAST_DB/htgs" -i ../9x12 -o 9x12.htm -T T
cd ..
$BLAST_CLI_HOME/blastall -a 2 -p blastx -e 1 -b 5 -v 5 -d $BLAST_DB/nr -i ../16x12 -o 16x12.htm -T T
perl /uufs/sequence/sys/blast/tblast_single.pl 16x12
cd TBLASTX
$BLAST_CLI_HOME/blastall -a 2 -p tblastx -e 1 -b 5 -v 5 -d "$BLAST_DB/nt $BLAST_DB/stds $BLAST_DB/est
$BLAST_DB/gss $BLAST_DB/htgs" -i ../16x12 -o 16x12.htm -T T
cd ..

```

FIG.15B

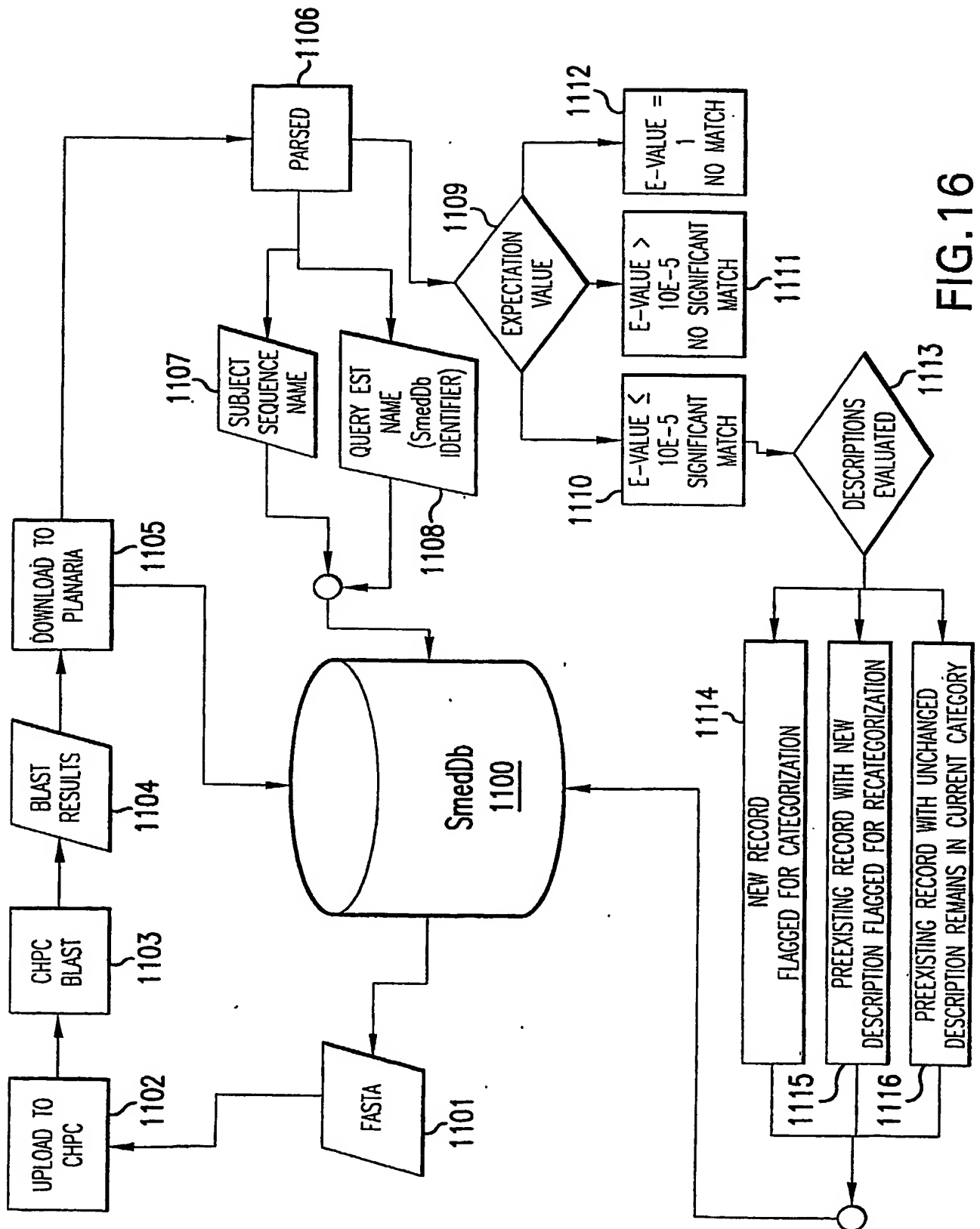


FIG.16

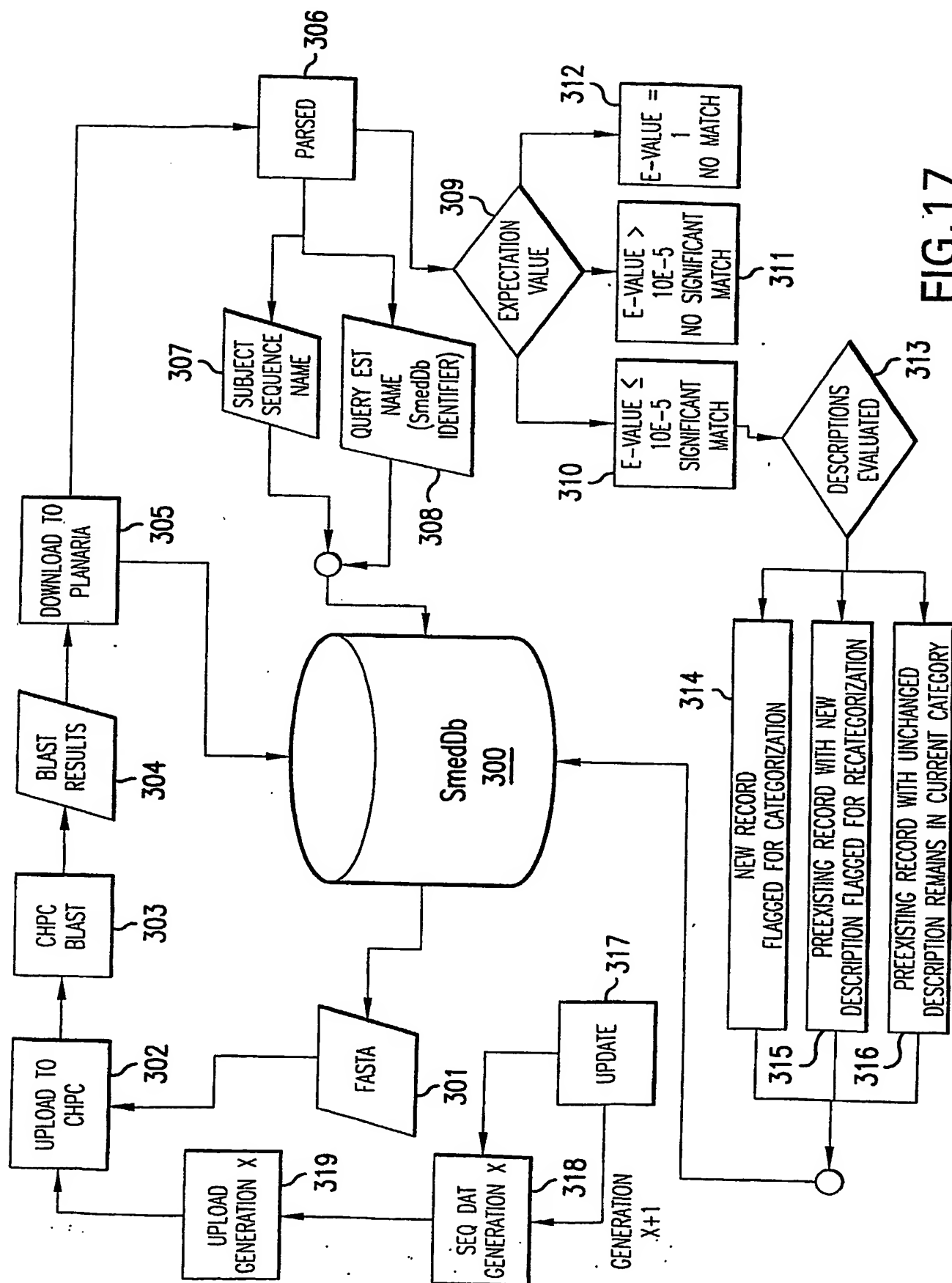
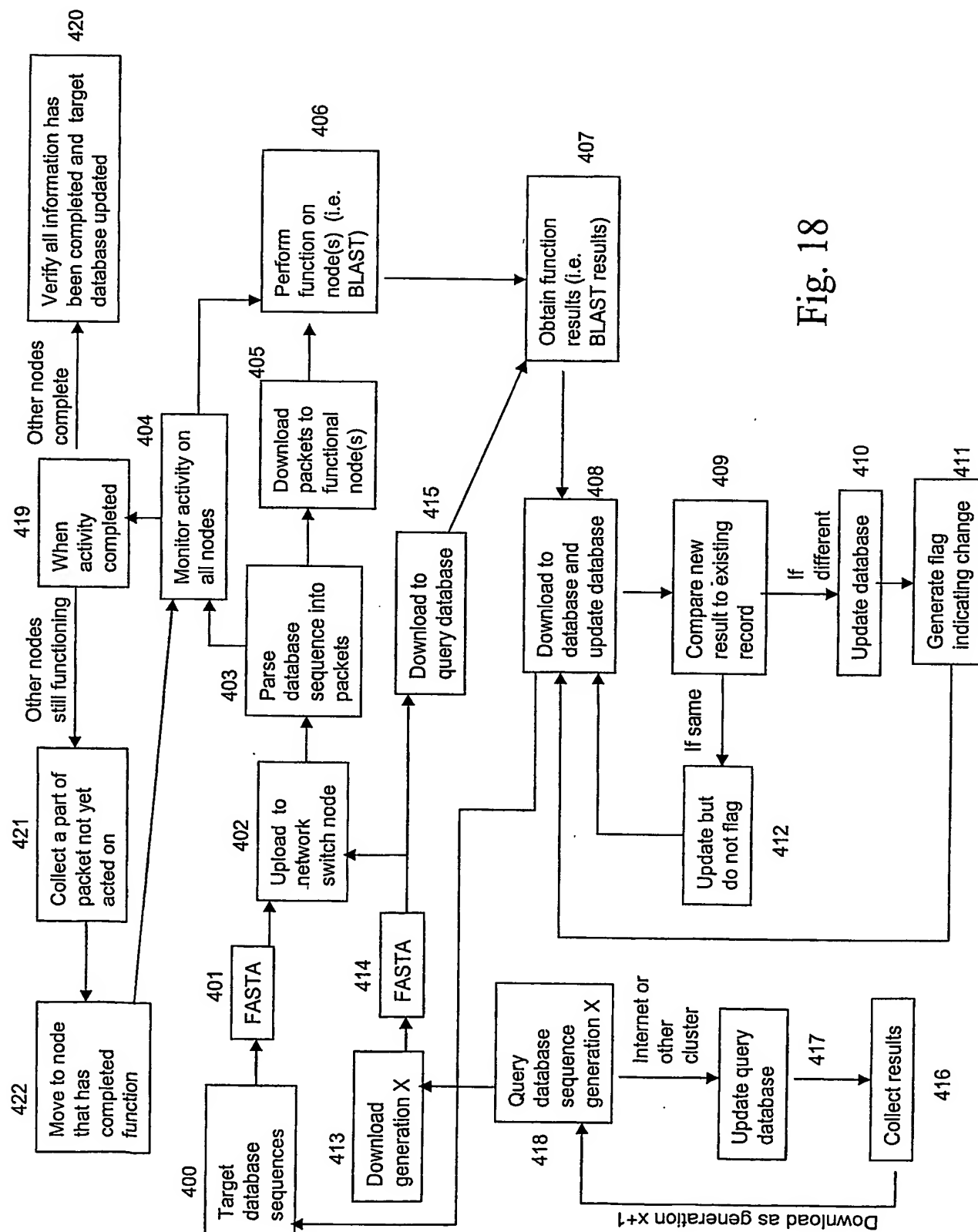


FIG. 17



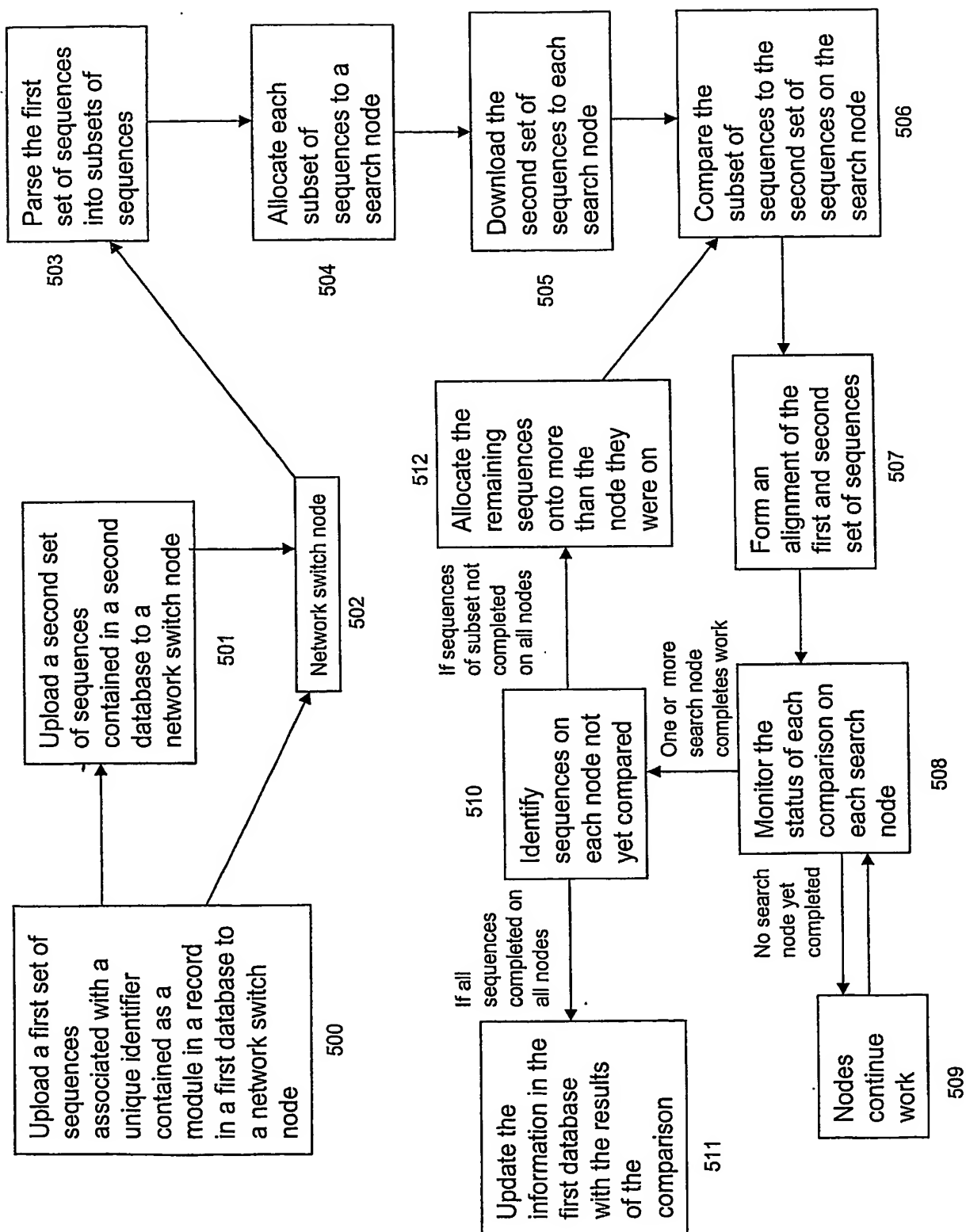


Fig. 19

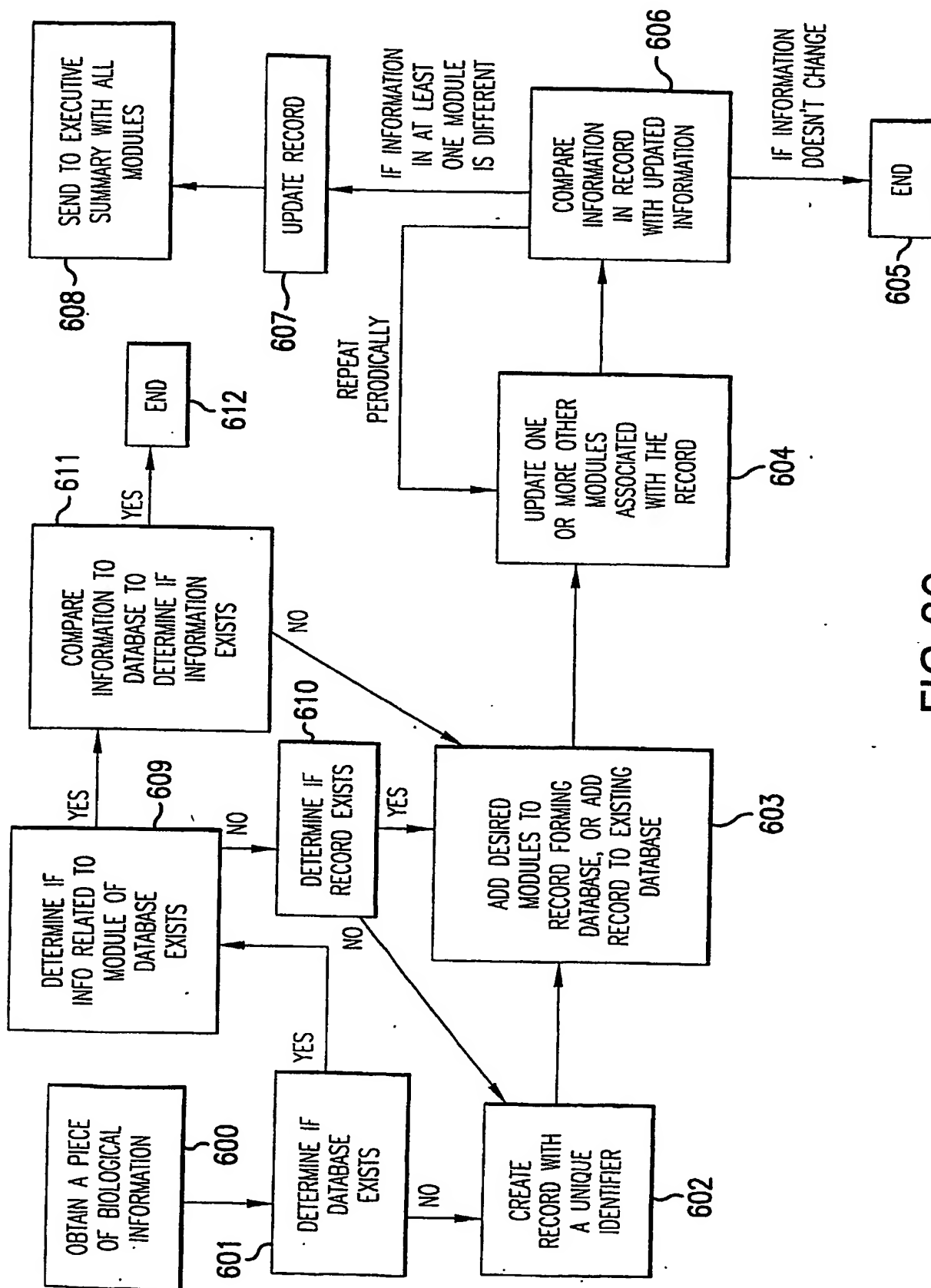


FIG. 20

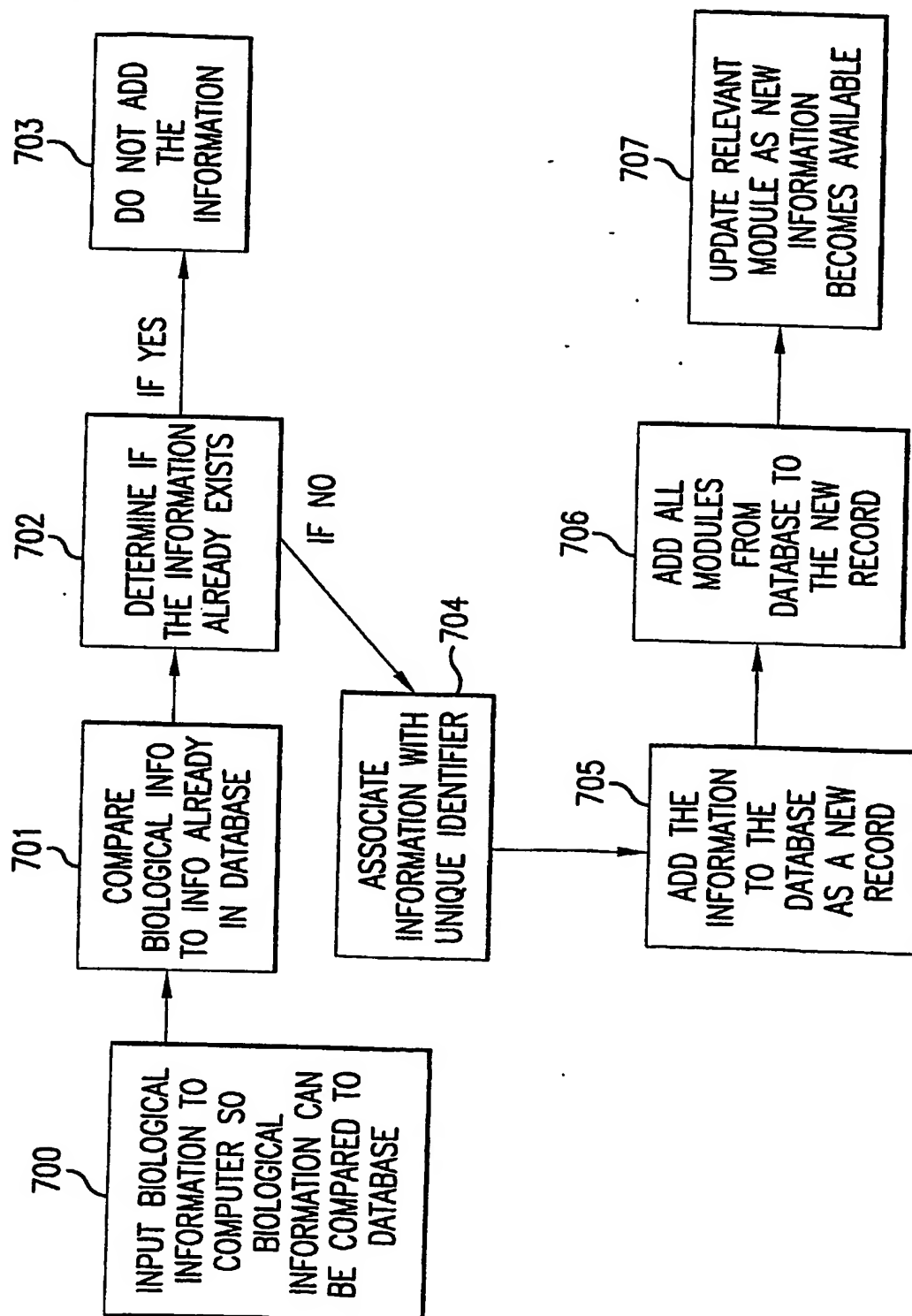


FIG. 21

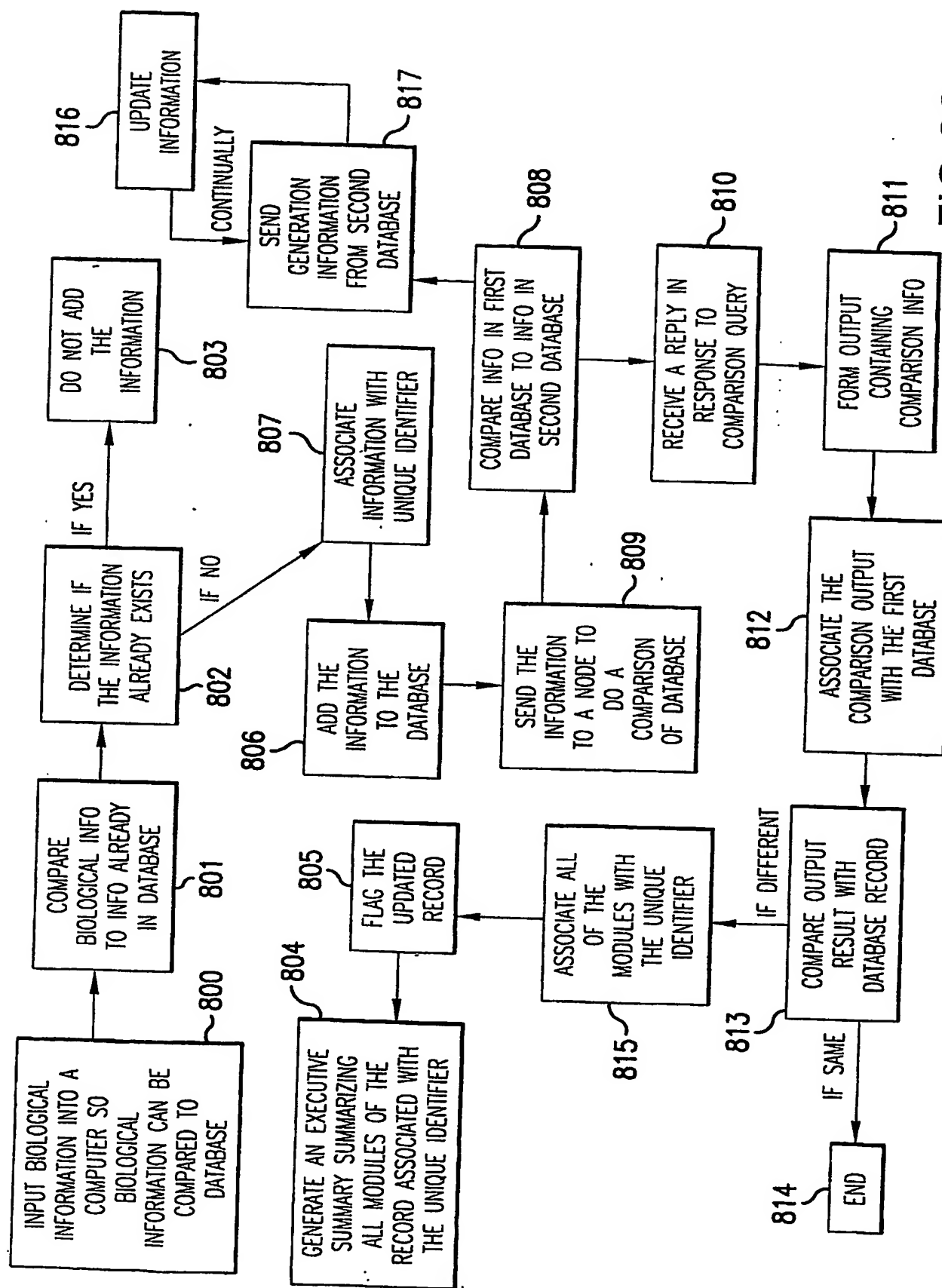


FIG. 22

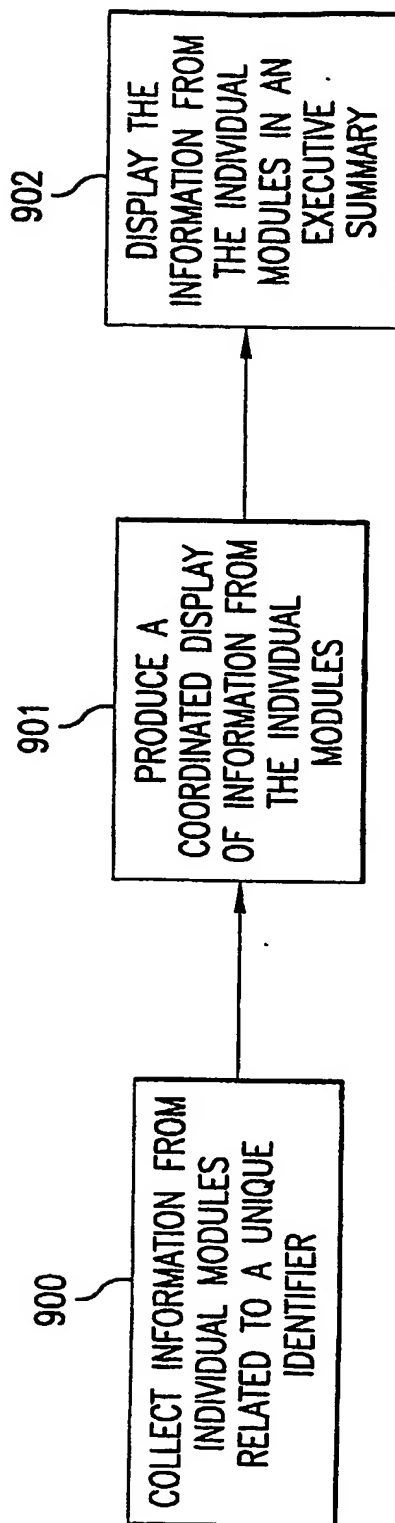


FIG. 23

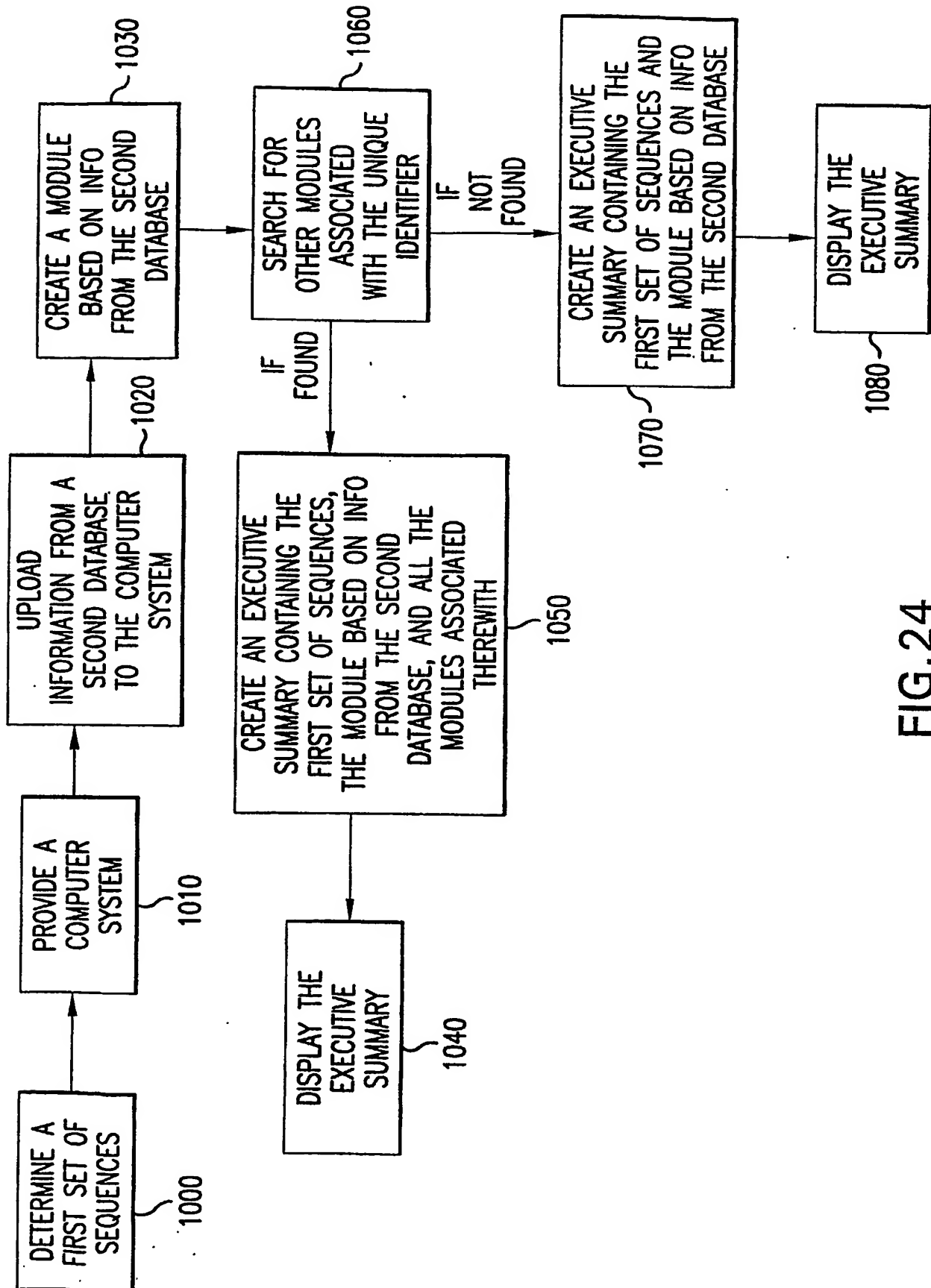


FIG. 24